REVIEW ARTICLE

ADHYAYAN Volume 14, Issue 2, 2024 Print ISSN: 2249-1066 Online ISSN: 2455-8656

Green AI: Minimizing Environmental Cost of AI Model Training and Deployment

Ankush Sharma

Software Architect, San Jose, California, USA.

Abstract

The rapid development of artificial intelligence (AI), particularly deep learning models, has contributed to transformative innovations across various industries. The environmental influence of AI model training and deployment, especially energy consumption and carbon emissions through large-scale computational tasks, has gained increasing attention. This paper explores the concept of "Green AI," a framework that emphasises minimizing the environmental costs of AI without sacrificing performance. By examining current practices in model development, energy consumption during training, and the role of sustainable deployment strategies, this research highlights practical solutions to mitigate AI's environmental footprint while encouraging more efficient and eco-friendly models.

Keywords: Green AI, sustainability, AI model training, environmental impact, energy efficiency, deep learning, carbon emissions, AI deployment.

Adhyayan: A Journal of Management Sciences (2024); DOI: 10.21567/adhyayan.v14i2.06

Introduction

Artificial intelligence (AI) has become a cornerstone of technological innovation, powering applications in healthcare, finance, manufacturing, and beyond (Sharma; 2025). As AI models, especially deep learning models, have grown in size and complexity, so too have the computational resources required for their development and deployment. These models require immense amounts of energy, leading to a corresponding increase in carbon emissions and environmental costs.

The environmental impact of AI model training and deployment is a growing concern. In 2019, it was reported that training a single large AI model could emit as much carbon as five cars over their entire lifetimes (Strubell *et al.*, 2019). This startling figure has sparked interest in the concept of "Green AI," a movement that seeks to make AI more environmentally sustainable. This study explains the energy-intensive nature of AI model training, the environmental costs associated with model deployment, and the strategies that can be implemented to minimise the environmental footprint of AI.

The Environmental Cost of AI Model Training

Energy Consumption in Model Training

Deep learning models are notoriously energy-hungry. State-of-the-art models, such as GPT-3, require billions of parameters and extensive computational resources to train. Training these models typically involves high-performance GPUs or TPUs running for days or even weeks at a time. According to recent estimates, training large models can consume thousands of kilowatt-hours (kWh) of electricity, often powered by non-renewable energy sources, leading to significant carbon emissions.

The energy consumption associated with model training varies depending on factors such as the size of the model, the architecture, and the hardware used. For instance, BERT, a popular natural language processing (NLP) model, consumed approximately 1,500 kWh during training (Brown *et al.,* 2020). When scaled to larger models, the energy usage increases exponentially, raising concerns about the long-term sustainability of this approach to AI development.

Corresponding Author: Ankush Sharma, Software Architect, San Jose, California, USA, e-mail: ankush.sh@ieee.org

How to cite this article: Sharma, A. (2024). Green AI: Minimizing Environmental Cost of AI Model Training and Deployment. Adhyayan: A Journal of Management Sciences, 14(2):28-30.

Source of support: Nil

Conflict of interest: None

Carbon Footprint and AI

The carbon footprint of AI models is primarily driven by the energy required to train them. Researchers have begun to quantify the carbon emissions linked with various AI models, revealing that training large models can generate significant amounts of CO2. A study conducted by Strubell *et al.* (2019) found that training a large NLP model could emit up to 284,000 kg of CO2, equivalent to the emissions from five cars over their lifetimes.

While these figures are alarming, they underscore the urgent need for more energy-efficient AI models and training processes. As AI continues to scale, developers and researchers must consider the environmental cost of training and take steps to mitigate their impact.

Toward Green AI

Model Efficiency and Optimization

One primary way to reduce AI's environmental impact is through model optimization. Techniques such as model pruning, quantization, and distillation allow for more efficient models that require fewer computational resources without sacrificing accuracy. For instance, model pruning involves eliminating unnecessary parameters from a neural network, reducing the model size and energy needed for training.

Another promising approach is advancing more efficient architectures. Researchers are exploring alternatives to the energy-intensive deep learning models that dominate AI research today. Sparse neural networks and low-rank approximations are two examples of architectures that can reduce the number of parameters and operations required for training, thereby lowering energy consumption.

Energy-Efficient Hardware

Further to optimizing AI models, using energy-efficient hardware can also have a key role in decreasing the environmental footprint of AI. Specialized AI accelerators, such as Google's Tensor Processing Units (TPUs), have been designed to provide high computational power while minimizing energy consumption. These devices can perform AI-related tasks more efficiently than traditional GPUs or CPUs, resulting in lower energy usage during training and inference.

Efforts are also being made to design hardware that is specifically tailored to the needs of AI models, balancing performance with energy efficiency. By leveraging energy-efficient hardware and designing models that can take full advantage of these devices, researchers can significantly decrease the environmental cost of AI development.

Sustainable AI Deployment

Cloud Computing and AI Deployment

Once trained, AI models are often deployed in cloud environments to support real-time applications, further contributing to energy consumption. Cloud-based AI services offer scalability and flexibility along with high energy usage, especially if the underlying infrastructure is powered by non-renewable energy sources.

Green cloud computing, which emphasizes employing renewable energy and better resource management, can mitigate some environmental impacts of AI deployment. Major cloud providers, such as Amazon Web Services (AWS), Google Cloud, and Microsoft Azure, have begun to invest in renewable energy and are aiming to power their data centres with 100% renewable energy in the coming years (Sharma; 2024). By selecting cloud providers prioritizing sustainability, AI developers can reduce the carbon footprint of their deployed models.

Edge Computing as a Solution

Edge computing offers another potential solution for reducing the environmental impact of AI deployment. By enabling data on local devices such as smartphones, IoT devices, or edge servers, AI models can diminish constant communication requirements with cloud servers, ultimately lowering energy consumption. Edge AI, which brings AI processing closer to the source of data, not only reduces latency and bandwidth usage but also minimizes the energy required for inference.

While edge computing may not completely replace cloud-based AI services, it provides a complementary approach that can lead to more sustainable AI deployment strategies (Sharma & Sharma, 2024).

Policy and Regulation

The Role of Policy in Promoting Green AI

Governments and regulatory bodies have an important role to encourage to develop the environmentally sustainable AI technologies. The governments can encourage AI developers to adopt greener practices by introducing policies supporting energy efficiency, carbon reduction, and responsible AI development,

For instance, Carbon pricing could be used as a tool to make developers and companies more conscious of their environmental impact. Additionally, research funding aimed at Green AI initiatives could help accelerate more efficient models and hardware.

International Cooperation on AI Sustainability

As AI development becomes a global endeavor, international cooperation will be critical to address the environmental challenges it poses. Collaborative efforts among nations, research bodies, and private enterprises can lead to sharing the best practices and developing global standards for AI sustainability.

The establishment of international frameworks for measuring and reporting the environmental impact of AI models would also help provide transparency and accountability, encouraging the adoption of Green AI principles worldwide.

Conclusion

As AI continues to grow in influence and importance, so too does the need for sustainable practices in its development and deployment. Green AI presents an opportunity to reduce the environmental costs associated with training and deploying AI models, without compromising on performance. By optimizing models, leveraging energy-efficient hardware, and adopting sustainable deployment strategies, the AI community can make meaningful progress toward reducing its carbon footprint.

In addition to technological solutions, policy and international cooperation will play a key role in promoting the adoption of Green AI. By working together, researchers, developers, policymakers, and companies can ensure that AI development is not only transformative but also environmentally responsible.

References

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, & Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877-1901.

Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. *arXiv preprint arXiv:1906.02243*.

Sharma, A., (2024) Chaos Engineering in Large Language Models: Resilience and Robustness Testing. Vivekananda Journal of Research [in press]; Accepted in November 2024

Sharma, A., Sharma, N., (2024) The Role of Artificial Intelligence in Revolutionising Financial Services: From Fraud Detection to Personalized Banking. Vivekananda Journal of Research [in press]. Accepted in December 2024.

Sharma, A. (2025). Applications of Generative AI in Healthcare: Transforming Medical Research, Documentation, and Patient Engagement. Global South Healthcare Journal, 1(1), 40–43.